Zeitschrift: Mémoires de la Société Vaudoise des Sciences Naturelles
Herausgeber: Société Vaudoise des Sciences Naturelles
Band: 17 (1978-1987)
Heft: 1

Artikel: L'information vue par un ingénieur
Autor: Coulon, Frédéric de
DOI: https://doi.org/10.5169/seals-259564

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 15.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

L'information vue par un ingénieur

PAR

FRÉDÉRIC DE COULON *

Résumé – *Pour l'ingénieur, l'information se présente sous deux aspects: celui, formel, du signal physique qui convoie l'information de sa source au destinataire – homme ou machine – et celui plus abstrait, parce que de nature statistique, de l'information effective transportée par ce signal. Comme mesure de cette information effective, on retrouve ici la notion d'entropie déjà rencontrée – mais dans un autre contexte – en thermodynamique.*

Le débit effectif d'information est souvent plus faible que le débit apparent associé au signal, ce qui caractérise la présence de redondances. Celles-ci peuvent être réduites par des procédures de codage appropriées.

Lors d'une transmission, les perturbations éventuelles (bruit) provoquent une modification du signal entraînant des erreurs d'interprétation. Ceci introduit une certaine ambiguïté sur la nature de l'information transmise qui en réduit le débit effectif. On appelle capacité d'une voie de transmission le débit effectif maximum qu'elle autorise pour un régime de perturbations donné. Sous réserve d'utiliser un débit d'information inférieur à cette capacité, l'introduction, par codage, d'une redondance structurée permet d'obtenir une diminution importante du taux d'erreurs sans réduction excessive du débit d'information.

1. INTRODUCTION

Le concept d'information peut être envisagé à différents niveaux:

1) formel,
2) technico-mathématique,
3) sémantique,
4) psychologique,
5) philosophique.

* Ecole Polytechnique Fédérale de Lausanne

Le point de vue de l'ingénieur n'intervient qu'aux deux premiers niveaux. Le niveau sémantique, qui touche à la signification d'un message, le niveau psychologique, qui concerne l'effet émotionnel (donc subjectif) de l'information, et le niveau philosophique, qui a trait à la connaissance et à l'origine de l'intelligence, ne sont pas interprétables en termes techniques.

Le niveau formel ne prend en compte que la nature du support de l'information, donc son *apparence.* Pour l'ingénieur, cette apparence est celle d'un *signal* physique transportant un flux continu ou discret de messages. Pour pouvoir aisément traiter (transmettre, transformer, interpréter) ce flux d'informations brutes, ce signal est avantageusement traduit sous forme électrique et représenté par des combinaisons de deux symboles élémentaires (0 et 1: signal numérique binaire). Le *débit apparent* d'une source d'information est ainsi exprimé en terme de symboles binaires (bits) par seconde.

Le niveau technico-mathématique est celui où l'on s'intéresse à la mesure du *débit effectif* de l'information convoyée par le signal et aux moyens possibles d'améliorer l'efficacité (recherche d'une plus grande concision) et la sécurité (lutte contre les erreurs) d'un transfert ou d'un stockage d'informations. Ceci nécessite le développement d'un modèle statistique de l'information effective (objective).

Ce sont de solides motivations techniques qui ont, à l'origine, conduit au développement de la *théorie du signal et de l'information,* appelée aussi théorie de la (ou des) communication(s). Les premières recherches ont été, en effet, motivées par le désir d'augmenter les cadences de transmission sur les lignes télégraphiques.

On s'est intéressé très tôt à la recherche d'une définition objective de l'information, capable d'en fournir une mesure quantitative. Les premières études sont dues à Nyquist et surtout à Hartley qui, en 1927, a montré dans un congrès international de téléphonie et de télégraphie comment il était possible de comparer les performances des divers systèmes de télécommunications. Il faudra toutefois attendre la publication des travaux de Claude SHANNON (1948), pour que les fondements de la théorie moderne de l'information soient fermement établis. Ces travaux, de même que l'ouvrage de Norbert WIENER (1948) «Cybernetics», qui aborde de manière parallèle les problèmes de la communication et du réglage automatique, ont eu une influence déterminante sur l'art de l'ingénieur des transmissions et du traitement de l'information.

2. SIGNAL ET INFORMATION

Les flux d'informations sur lesquels l'ingénieur désire pouvoir agir proviennent de nombreuses sources et concernent principalement l'observation (mesure) de phénomènes physiques ou le dialogue (communication)

entre hommes, entre l'homme et la machine, ou entre les machines elles-mêmes.

Pour pouvoir aborder le processus de transfert de l'information en termes généraux, il est utile de se référer au modèle présenté sur la figure 1.

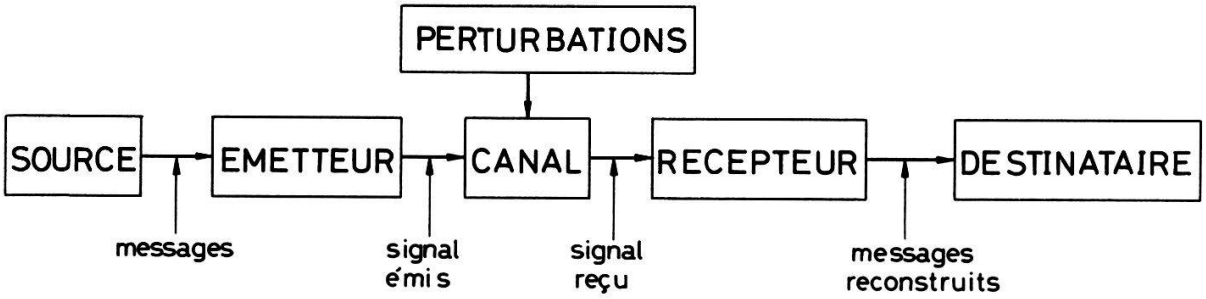

Figure 1. – Système de transfert d'informations.

La source d'information sélectionne les messages. Ceux-ci peuvent consister en mots écrits ou parlés, en chiffres, en images, en musique, etc. L'émetteur transforme le message en un signal, généralement électrique, qui est alors envoyé par une voie de communication (canal) de l'émetteur au récepteur. Celui-ci retraduit le signal reçu en un message qui est fourni au destinataire. Le signal, lors de sa transmission, est toujours plus ou moins altéré par la présence inévitable de phénomènes perturbateurs (bruit de fond, distorsions, etc.).

Le signal est donc le support physique de l'information, son apparence en quelque sorte. Il représente ce que nous conviendrons d'appeler l'information brute. Il peut être de nature diverse: optique, acoustique, électrique, magnétique, mécanique, thermique, etc. Grâce à des capteurs, ou transducteurs, appropriés (par exemple: cellule photo-électrique, microphone, élément piézo-électrique, thermocouple, etc.) ce signal est généralement traduit sous forme électrique afin de tirer parti des immenses possibilités offertes par les systèmes électroniques de transmission et de traitement de l'information. Une conversion inverse est souvent réalisée au niveau du récepteur de manière à restituer l'information au destinataire sous sa forme originale.

3. Mesure de l'information apparente

La source produit soit un flux d'informations discret, soit un flux d'informations continu. Le signal qui lui sert de support est donc, par nature, lui aussi discret ou continu.

Le premier cas correspond à la génération d'une séquence de *signes* conventionnels distincts, tels que lettres, chiffres, symboles graphiques, impulsions lumineuses, etc. C'est la situation rencontrée dans les transmissions télégraphiques, le télex, ou la transmission de données entre machines (ordinateurs et équipements périphériques).

Dans le second cas, le message est représenté par une fonction continue du temps traduisant l'évolution chronologique d'une grandeur physique. Un exemple en sont les fluctuations de la pression acoustique due à la voix ou à un instrument de musique.

Comment mesurer le flux d'information brute, c'est-à-dire le débit apparent d'information?

Pour un flux discret, il suffit de compter le nombre de signes – lettres ou chiffres par exemple – apparaissant dans un intervalle de temps prescrit. Toutefois, pour pouvoir comparer des flux d'informations provenant de sources de nature différente, il est nécessaire de pouvoir se baser sur une unité de comptage commune.

Lorsqu'une personne parle à une autre, le langage est le véhicule et le code par lequel l'information est transmise. Les mots en sont les constituants de base. Or le nombre de mots d'un grand dictionnaire est considérable (plusieurs centaines de milliers), même si nous ne les utilisons pas tous dans la langue courante. Pour les communications écrites, nous utilisons les 26 lettres de l'alphabet, plus un 27^{e} qui est l'espace, comme constituants de base. Remarquons que, dans ce cas, le nombre de constituants est déjà relativement petit.

Ceci nous conduit tout naturellement à la question: y a-t-il une forme de représentation de l'information basée sur un dictionnaire minimum? Ou, en d'autres termes, peut-on définir une unité – quantum élémentaire – d'information?

La response est évidemment oui. Car il suffit, en effet, d'un «alphabet» élémentaire de deux symboles pour que, par combinaison de L d'entre eux, il soit possible de représenter (coder) $n = 2^{L}$ états distincts d'un alphabet plus riche. Ceci correspond à une transcription du message initial en une combinaison de deux déclarations élémentaires, s'excluant mutuellement, telles que oui/non, ouvert/fermé, blanc/noir, pile/face, etc. Il est devenu usuel de représenter les deux choix offerts par les chiffres 0 et 1.

On observera ici que, d'une manière générale, le nombre de signes différents pouvant être exprimés à l'aide d'une séquence de symboles élémentaires de longueur L croît exponentiellement avec L. Inversement, L croît comme le logarithme du nombre de signes n de l'alphabet (on utilisera ici pour indiquer le logarithme de base 2 la notation conventionnelle: $\mathrm{lb}\, x = \log_2 x$)

$$\mathrm{L} = \mathrm{lb}\, n \qquad (1)$$

Ainsi, la longueur L de la séquence binaire d'encodage fournit une mesure de la quantité d'information apparente (appelée parfois «quantité de décision») associée à chaque signe du message. Nous nous conformerons, dans ce qui suit, aux recommandations de l'Union internationale des télécommunications (UIT) en appelant «bit» l'élément binaire, 0 ou 1, employé pour le codage de l'information. Ce terme, d'origine américaine, fut initialement proposé pour caractériser l'unité binaire d'information effective (voir § 4). Toutefois, l'usage l'a aujourd'hui consacré comme unité binaire d'information apparente. Chaque signe d'un alphabet donné contient donc L bits.

Par exemple, les 26 lettres de l'alphabet ordinaire peuvent être représentées par des blocs de cinq bits ($2^5 = 32$). C'est la solution exploitée par le code Baudot utilisé pour les transmissions télégraphiques internationales:

A ⟷ 0 0 1 1 1
B ⟷ 0 1 1 0 0

C ⟷ 1 0 0 0 1
D ⟷ 0 1 1 0 1
etc.

Afin de permettre la transmission non seulement de lettres, mais aussi des chiffres 0 à 9, ainsi que de certains signes spéciaux, chaque groupe de 5 bits est en fait utilisé dans deux acceptions différentes, ce qui nécessite une double table de codage. Le choix de la table utilisée est alors indiqué en intercalant dans le flot d'informations un signe particulier. Si la source produit une séquence de signes à la cadence de 10 signes par seconde, on obtient un débit apparent (aussi appelé «débit littéral»)

$$D = 10 \times 5 = 50 \text{ bits/s.}$$

Un autre exemple est le codage alphanumérique utilisé pour la transmission de données et d'ordres dans les systèmes informatiques (code ISO ou USASCII) qui évite la double signification attribuée à chaque mot dans le code Baudot. L'emploi de mots de 7 bits b_7, b_6, ..., b_1 permet de représenter théoriquement $2^7 = 128$ lettres, chiffres, signes spéciaux ou ordres différents. La figure 2 présente le code ISO (International Standardization Organization) à 7 bits. Les différents messages y sont regroupés dans une matrice de 16 lignes et 8 colonnes ($8 \times 16 = 128$). Chaque ligne et chaque colonne est identifiée par un groupe de 4, respectivement 3, symboles binaires.

b_7	b_6	b_5	b_4	b_3	b_2	b_1		0	0	0	0	1	1	1	1
								0	0	1	1	0	0	1	1
								0	1	0	1	0	1	0	1
								0	1	2	3	4	5	6	7
0	0	0	0				0	NUL	DLE	Space	0	—	P	@	p
0	0	0	1				1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0				2	STX	DC2	"	2	B	R	b	r
0	0	1	1				3	ETX	DC3	#	3	C	S	c	s
0	1	0	0				4	EOT	DC4	$	4	D	T	d	t
0	1	0	1				5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0				6	ACK	SYN	&	6	F	V	f	v
0	1	1	1				7	Bell	ETB	'	7	G	W	g	w
1	0	0	0				8	BS	Cancel	(	8	H	X	h	x
1	0	0	1				9	HT	EM	)	9	I	Y	i	y
1	0	1	0				10	LF	SS	*	:	J	Z	j	z
1	0	1	1				11	VT	Escape	+	;	K	[	k	
1	1	0	0				12	FF	FS	,	<	L	\	l	
1	1	0	1				13	CR	GS	-	=	M	]	m	
1	1	1	0				14	SO	RS	.	>	N	^	n	
1	1	1	1				15	SI	US	/	?	O	_	o	Delete

Figure 2. – Code ISO à 7 bits.

Ainsi la lettre A ou l'ordre CR *(carriage return),* qui permet de commander le retour du chariot d'un télétype, correspondent aux mots binaires respectifs:

$$A \longleftrightarrow 1000001$$
$$CR \longleftrightarrow 0001101$$

Une représentation sous forme binaire de l'information est évidemment très favorable du point de vue technique. Les symboles de codage 0 et 1 peuvent en effet facilement correspondre à deux états stables d'un dispositif électrique simple tels que l'ouverture et la fermeture d'un contact (interrupteur électronique ou mécanique) entraînant le passage ou l'absence de courant. Cette succession de présences ou d'absences de courant est aisément transmise et détectée par le système destinataire. La figure 3 illustre le codage binaire d'un message primaire, traduit à son tour en un signal représentatif où le symbole 1 correspond à la présence d'une impulsion de courant dans un intervalle de temps prescrit T, et le symbole 0 à l'absence d'une telle impulsion. On qualifie un tel signal de «numérique» (ou digital).

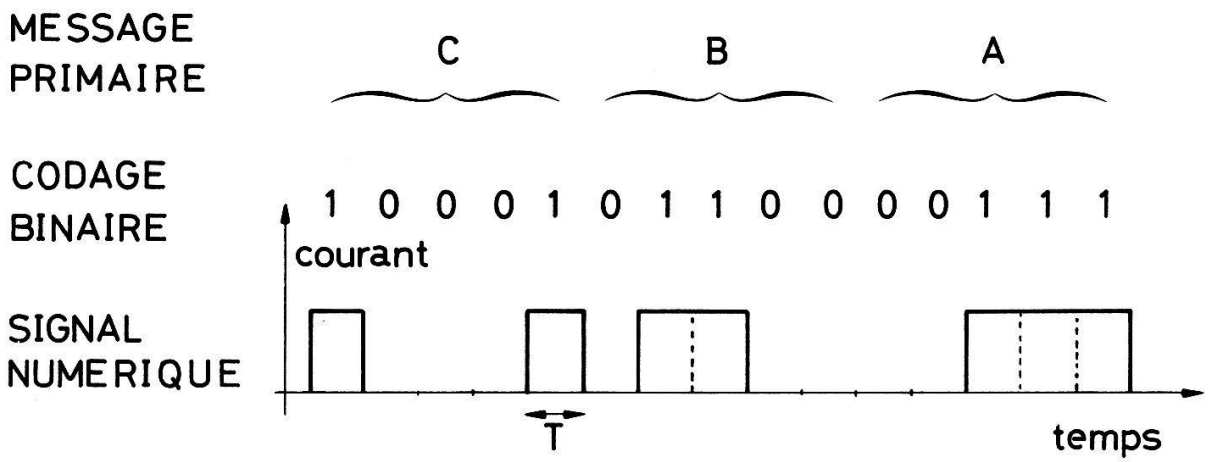

Figure 3. – Traduction d'un message codé en un signal représentatif.

Puisque les nombres décimaux sont également représentables sous cette forme, un traitement de l'information est possible à la condition de transposer les opérations mathématiques et logiques usuelles (addition, multiplication, plus grand que, plus petit que, etc.) dans un système binaire. C'est le mode de fonctionnement des calculatrices électroniques dont on connaît l'essor prodigieux.

Peut-on également appliquer cette unité d'information apparente, de nature fondamentalement discrète, pour mesurer un flux d'information continu? Bien que cela puisse paraître étrange, la réponse est encore oui! Pour le montrer, il faut disposer d'un principe de quantification qui nécessite l'introduction d'un concept nouveau: celui d'échantillonnage.

On sait qu'un flux continu d'informations est représenté par une fonction continue du temps. Celle-ci se traduit sous la forme d'un signal, dit «analogique», dont l'amplitude peut théoriquement prendre une infinité de valeurs différentes les unes des autres. C'est le cas du courant microphonique traduisant, sous forme électrique, les ondes acoustiques de la parole, par exemple. Toutefois, il est expérimentalement impossible de mesurer (observer) avec exactitude chacune de ces valeurs d'amplitude. C'est pourquoi on peut faire correspondre à cet ensemble continu de valeurs, connues avec une précision limitée, un nouvel ensemble discret (dénombrable) de valeurs distinctes garantissant la même précision. C'est une procédure de quantification en n niveaux distincts, dont les amplitudes sont souvent un multiple d'un quantum q qui correspond au plus petit écart mesurable.

Shannon a, par ailleurs, démontré qu'il n'est nullement besoin de connaître en tout instant t la valeur prise par un signal analogique. Si l'on analyse la façon dont un tel signal varie avec le temps, on constate qu'il comprend des composantes de beaucoup de fréquences différentes. Celles-ci occupent un certain domaine de fréquence que l'on appelle la largeur de bande B du signal. Par exemple, on sait que les sons produits par des

instruments de musique ont des composantes de fréquence allant de quelques dizaines de hertz (le hertz est une unité de mesure de fréquence égale à un cycle par seconde) à plusieurs milliers de hertz. Une largeur de bande de l'ordre de 15 à 20 kHz est exigée des installations d'amplification et de reproduction électro-acoustiques dites «à haute fidélité». Or, on constate qu'il est théoriquement possible de reconstruire entièrement un signal analogique à partir de prélèvements ponctuels – appelés échantillons – de sa valeur instantanée effectués à intervalles de temps réguliers T_e, à la condition que

$$T_e \leq \frac{1}{2B} \tag{2}$$

Autrement dit, l'échantillonnage doit être effectué à une cadence supérieure ou égale au double de la plus haute fréquence contenue dans le «spectre» (distribution de l'énergie en fonction de la fréquence) du signal. La reconstruction est obtenue en multipliant chaque échantillon par une fonction d'interpolation du type $\sin x/x$, avec $x = 2\pi Bt$, et en effectuant la somme.

Ce résultat est le *théorème d'échantillonnage* qui joue un rôle fondamental en théorie du signal. Soit n le nombre de niveaux de quantification utilisables, ordonnés de zéro à $n - 1$: chaque niveau est alors identifiable par son numéro d'ordre. Celui-ci peut être exprimé à son tour sous la forme d'un nombre binaire de L bits (avec L le plus petit entier supérieur ou égal à lb n). Le principe de l'échantillonnage, de la quantification et du codage binaire d'un signal analogique est illustré sur la figure 4. →

Il en découle qu'un segment de signal analogique de durée T peut être représenté par un nombre d'échantillons

$$N = 2BT \tag{3}$$

ou un nombre de symboles binaires

$$NL = 2BTL \text{ bits} \tag{4}$$

En d'autres termes, le flux d'information brute associé à un tel signal possède un débit apparent

$$D = 2BL \text{ bit/s} \tag{5}$$

Par exemple, dans le cas d'un signal téléphonique, la largeur de bande B est de l'ordre de 4000 Hz et une quantification (non linéaire) avec $n = 2^8 = 256$ niveaux est nécessaire pour garantir une qualité suffisante de repro-

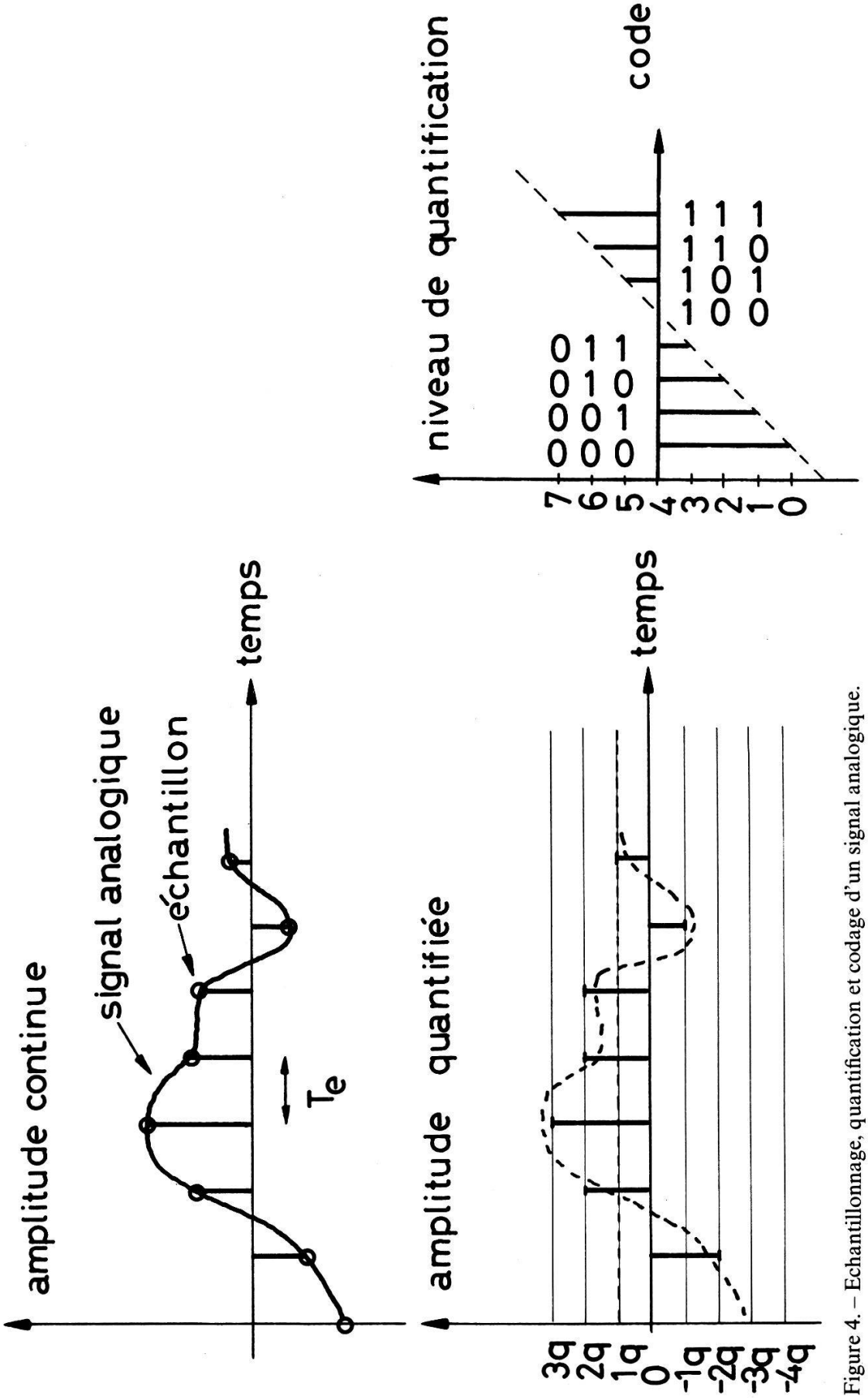

Figure 4. – Echantillonnage, quantification et codage d'un signal analogique.

duction. La transmission d'un tel signal sous forme numérique (système dit à modulation par impulsions codées ou PCM) entraîne un débit apparent d'information

$$D = 2B \cdot \text{lb}\, n = 64000 \text{ bit/s}.$$

4. Mesure de l'information effective

Considérons une source d'information qui émet des prévisions météorologiques. Parmi les messages possibles figurent les indications «pluie» et «neige». Ces deux messages contiennent le même nombre de lettres (cinq), dont chacune peut être codée par une séquence binaire distincte de 5 bits, de sorte qu'ils représentent la même quantité d'information apparente (25 bits).

Sous nos climats, au mois d'août par exemple, la pluie est un phénomène relativement fréquent, alors que l'apparition de neige à basse altitude est tout à fait exceptionnelle, voire rarissime. On réalise donc que chacun de ces deux messages ne transporte pas la même quantité d'information effective. Celle-ci dépend en quelque sorte du caractère imprévisible, incertain, de l'événement. Un message annonçant un événement certain n'apporte aucune information effective, puisqu'il n'est même pas nécessaire de le recevoir pour avoir connaissance de l'événement. Inversement, la réalisation d'un événement très improbable est accompagnée d'un effet de surprise considérable, et l'information effective acquise est très grande. On peut donc dire qu'un message fournit d'autant plus d'information qu'il est original.

Il convient de souligner que le sens que nous donnons ici à «information effective» est dénué de tout aspect sémantique ou psychologique. Il correspond simplement à une mesure objective de la liberté de choix dont on dispose lorsque l'on sélectionne un message, et non à l'interprétation subjective de celui-ci.

La théorie de l'information, développée par Shannon, est une théorie stochastique des messages. Elle introduit une mesure statistique de la quantité d'information effective associée à chaque symbole produit par une source d'information donnée. Par extension, elle permet de déterminer la capacité – ou cadence maximale de transmission de l'information – d'un système, qu'il soit ou non soumis à un régime de perturbations aléatoires. Elle fournit avant tout une méthode d'estimation des limites de performance d'un «système informationnel» et des critères permettant la comparaison de différents systèmes.

Puisque la quantité d'information effective I_A associée à la réalisation d'un événement (ou symbole) A est une mesure de l'incertitude que l'on a

a priori à son sujet, il est naturel de l'exprimer mathématiquement comme une fonction de l'inverse de sa probabilité d'apparition p_A. Cette fonction doit être telle qu'à $p_A = 1$ (événement certain) correspond $I_A = 0$. Il est de plus raisonnable de supposer que la quantité d'information totale associée à une séquence de symboles indépendants soit égale à la somme de leurs quantités d'information individuelles. La relation fonctionnelle satisfaisant simultanément ces conditions est le logarithme, puisque $\log 1 = 0$ et $\log (A \cdot B) = \log A + \log B$.

Ainsi, la quantité d'information effective associée au symbole A, de probabilité p_A, est définie par

$$I_A = \log 1/p_A = -\log p_A \tag{6}$$

Il est commode d'utiliser dans cette définition un logarithme de base 2, de telle sorte que l'unité d'information effective corresponde à la quantité d'information associée à la réalisation d'un événement binaire à deux états *équiprobables.*

C'est cette unité d'information effective qui, à l'origine, a été appelée «bit». Toutefois, comme nous l'avons signalé plus haut, ce terme dénote aujourd'hui l'élément binaire (0 ou 1) servant d'unité de mesure de l'information apparente. Aussi, pour éviter toute confusion, nous nous conformerons ici aux recommandations de l'UIT en appelant l'unité d'information effective un «shannon», en hommage au père de la théorie de l'information.

Un shannon (en abrégé: sh) est donc la quantité d'information effective transportée par un bit pour autant que les symboles binaires soient équiprobables. Par contre, si la probabilité du symbole 0 vaut $p_0 = 0{,}2$ et celle du symbole 1 vaut $p_1 = 1 - p_0 = 0{,}8$, on obtient: $I_0 = -\text{lb}\ 0{,}2 = 2{,}322$ sh et $I_1 = -\text{lb}\ 0{,}8 = 0{,}322$ sh. La quantité d'information effective par symbole fournie *en moyenne* par la source est alors simplement calculée, si les symboles sont indépendants, en effectuant la somme des termes I_0 et I_1 pondérés par la fréquence relative (probabilité) d'apparition des symboles correspondants: $H = p_0 I_0 + p_1 I_1 = 0{,}722$ shannon par bit (espérance mathématique, ou moyenne statistique, de la quantité d'information associée à chaque symbole).

D'une manière générale, pour une source générant des messages construits à partir d'un alphabet de n signes différents et indépendants ayant les probabilités d'apparition respectives $p_1, p_2, \ldots, p_i, \ldots, p_n$, la quantité d'information effective moyenne fournie par chaque signe est donnée par la formule (l'indice 0 dénote ici une source *sans mémoire,* c'est-à-dire sans relations conditionnelles entre signes):

$$H_0 = -\sum_{i=1}^{n} p_i \cdot \text{lb}\, p_i \quad \text{(sh/signe)} \tag{7}$$

Cette quantité est appelée *l'entropie* de la source, terme choisi en raison de la similitude de la formule (7) avec celle de l'entropie thermodynamique. A titre d'exemple, la figure 5 représente la variation de l'entropie d'une source binaire sans mémoire en fonction de la probabilité d'un des symboles.

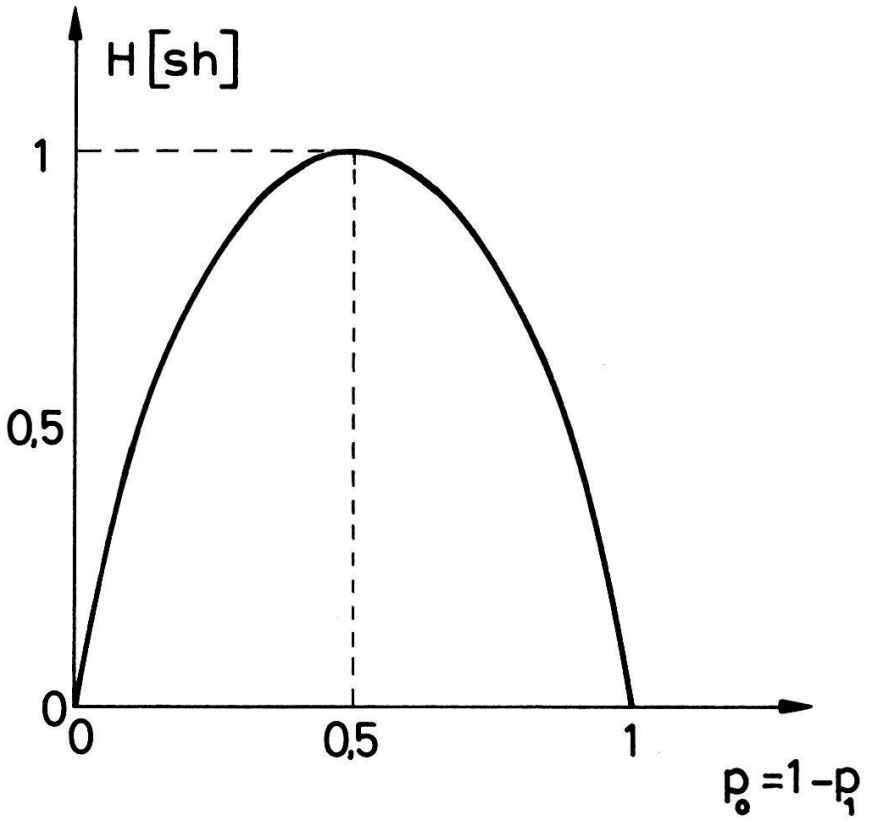

Figure 5. – Entropie d'une source binaire.

On démontre facilement que l'entropie atteint une valeur maximum lorsque tous les signes sont équiprobables:

$$H_{max} = \text{lb } n \text{ pour } p_i = 1/n \tag{8}$$

L'entropie maximum s'identifie donc avec la quantité d'information apparente – ou quantité de décision – exprimée en (1).

Des relations conditionnelles de signe à signe existent fréquemment. C'est, par exemple, le cas de la langue écrite où certaines lettres apparaissent plus fréquemment à la suite d'autres lettres. Ainsi la lettre q est toujours suivie d'un u en français, sauf quand elle est en position terminale. L'entropie de telles sources d'information peut être estimée en recourant à un modèle statistique plus élaboré (processus de Markov) faisant intervenir toutes les probabilités conditionnelles, dénotées $p_{i|k}$, pour qu'un signe i soit précédé par une séquence k donnée composée de m signes *(mémoire d'ordre*

m). Si la source dispose de n signes différents, on peut construire n^m séquences k différentes apparaissant avec la probabilité p_k. On obtient alors

$$H_m = \sum_{k=1}^{n^m} p_k \left[- \sum_{i=1}^{n} p_{i|k} \operatorname{lb} p_{i|k} \right] \tag{9}$$

Pour une source donnée, les entropies calculées sont d'autant plus faibles que le degré de mémoire pris en compte dans le modèle est élevé. La valeur asymptotique obtenue en faisant tendre m vers l'infini est l'entropie limite H_∞ qui mesure la quantité d'information effective moyenne produite par la source.

$$H_\infty \leq \ldots \leq H_m \leq \ldots \leq H_0 \leq H_{max} \tag{10}$$

Pour la langue française écrite, utilisant 27 signes différents (26 lettres + l'espace), $H_{max} = \operatorname{lb} 27 = 4{,}755$ sh/signe. Une source sans mémoire générant les signes de cet alphabet en respectant les probabilités d'apparition de ces signes possède une entropie $H_0 \simeq 3{,}9$ sh/signe. Si l'on considère un texte élaboré par une telle source, il n'a apparemment que peu de rapport avec la langue française. Une meilleure approximation est obtenue en introduisant un certain degré de mémoire. On a reproduit sur la figure 6, à titre d'illustration, deux exemples de simulation de la langue française réalisés dans le cadre d'un travail de diplôme (P.A. Rumley, diplôme d'ingénieur-électricien, EPFL 1971). Ils correspondent à la simulation d'une source ayant pour le premier une mémoire d'ordre 1 et pour le second une mémoire d'ordre 2. Seules des contraintes statistiques sont ici respectées, il n'y a bien entendu aucune contrainte grammaticale.

Pour une mémoire d'ordre 1, l'entropie H_1 tombe à environ 3,1 sh/signe. Elle continue à décroître lorsque le degré de mémoire augmente en se rapprochant asymptotiquement de l'entropie limite H_∞. Celle-ci a été estimée pour la langue écrite anglaise à une valeur voisine de 1 sh/signe. On peut présumer qu'il doit en être à peu près de même pour la langue française.

Si H est l'entropie d'une source et τ est la durée moyenne d'émission d'un signe, on peut déterminer un débit effectif d'information

$$H' = H/\tau \text{ (sh/s)} \tag{11}$$

a) Source avec mémoire d'ordre 1 :

R DE D CLE A E COTRINT ETE D BLU VOITRE S DE I A ME QUTAVEL AR CHE
MAILILGES T VE JES DEI LUGE JE QUER JETEMEMEN GESURE A E ASERE DE
DET ER MMBOUX DRESSERIN HEMMBTONTE S OMAGN NS VAR LUIT RMIS DEREURE
ENAUSTEURMA PE SAIR PL A NA IN MI N OITRE LELE BJEMOU CORE JEQU
NCUNTE QUIFA CEMES E PRIREUF TE PAMEN BES DUN DIT E DESAUI AITE
CEURILETONESUR PA CAVIS I MENDRADIRER S ASS LLOMEUEMAVE TRENEREAIOUN
JA JE IERTCISTE S E BREANT QUI LIS SET APLEN PITA E CE EUVUT HAIT
EPAND DENN C QUTR JA E ATOUS JE SUS Q URINASE SI S AVRE LUR AQUEPAISSULE
D DETE DUREMM D DERIT AIVRRET LL MER QULHIS INE VE S NON SUAIN QUTRISER
LLENLUERE LE CE PAMIRE SAM OUROUGRANN PAS ME ABOTE PERE LEN DENT JANTRES
UDEIE J DE HE S DENAN IT ENIR TARE MA DIEN DE LSENT LET PULU JE MERE
CE S E POUFUR AT HOR ANS FELAVULE LA S IAMARI J F PU IENN FRAISER RUE
HOMA LLARI MESOUN AIN NS NTAQUX ILOCEU E SS S NE L DE JELANGNS IT
QU JE PLI L UILU L LAOUTU A ANSE SUS CRINTA E NE HEASE LLE L A
DEUTEMANCEM MAFIQUCOIOUIRE CE QUILELE JERNN VIR N DAMOD

b) Source avec mémoire d'ordre 2 :

AIMME ET PEUX DESCACUGLEUX PLUS UN LA SOUTE DON TOMMENT IL RE AMIERMIS
Y ALMEN MONS PETENTS PASSE PR EAUR ILLOIRE RACIENTREVER SITES IMPARMA
ISSE CET IL ASSINUTOUT ENT RE COUT DU ILS MOI PAS DEUREZ FRANC HE UNE
TCU ONT CHEVRE ENT QU IL A SORTE ET UNE VEUX LA TOUSE VOIRANG IL TEMEM
A MOIS SEPASSENCE PUI ME ET A BRATICN RE LAIS VCIRESTEMPS UN TOUT CON
T MAIEN ETTE ETAIN ORS QUE BRE GRIFRME N CEVERS FEUS PAS DITU ON PAS A
POUVEC CRES REAU PERTAIS L HE COMMELLE AERRE IL HARRIRME TAIS OUPAS U
NE MA T QUE N JE TAGE QUI FINSA TOUE CHAQUISPCIR MATAIT L SUADMIT DERA
NCOUR DE DE COURES GRANTRES PIEN LONT ABLAIS DE ALE PET LA D A LOUT PR
CI ES ALLE MES ETAI LE JOUR SE MOITE JE SUIVINEABLET LES ACE DES DES L
UILAMARIS TCUVRE IDES IGIQUANDES QUE L SE SIL NOUS POURNANT MAIR ESCES
BUCHIPONS AVAIERSA VOUST JE DE LEMINEE VEN N Y PRE DE FAIT QUE DOURS I
L ENTEMONNE VAIR TONCE UN REPENTRE ACI DEUROYA A MOND UNE ILIT L OUDES
PARAIST DEVROIR CONS LE CLES PEUX DUNE PLA FONDS CET MEME BUS LILLES C
ES AUTEMPS SUCULANT D

Figure 6. – Simulation élémentaire de la langue française.

5. REDONDANCE DE L'INFORMATION

La différence entre l'entropie maximum (égale à la quantité d'information apparente) et l'entropie limite (correspondant á la quantité d'information effective) constitue la *redondance* de l'information

$$\Delta H = H_{max} - H_{\infty} \tag{12}$$

que l'on exprime généralement en valeur relative

$$r = \frac{\Delta H}{H_{max}} = 1 - \frac{H_{\infty}}{H_{max}} \tag{13}$$

En paraphrasant Shannon, on peut définir la redondance comme cette fraction du message qui n'est pas nécessaire, en ce sens que, si elle manquait, le message resterait complet ou du moins pourrait être complété.

Dans le cas de la langue écrite, cette redondance paraît être de l'ordre de 70 à 80%. Pour la langue parlée, on l'estime à environ 60%. Les images de télévision ont une redondance variant généralement entre 60 et 90%, selon leur richesse en détails. La redondance de la plupart des documents fac-simile tels que lettres commerciales, cartes météorologiques, dessins techniques, etc., se situe au-dessus de 80%.

La suppression, même partielle, de cette redondance avant transmission ou stockage améliore l'efficacité de la source d'information. Dans le cas d'un stockage, cela permet de diminuer la dimension de la mémoire nécessaire. Mais comment opérer cette réduction?

Une solution évidente, déjà exploitée dans son principe voilà plus d'un siècle par Morse pour son code télégraphique, consiste à encoder les signes fournis par la source à l'aide de mots binaires dont la longueur l_i (nombre de bits) varie en raison inverse de leur probabilité p_i d'apparition. La longueur moyenne de ces mots vaut alors

$$L = \sum_{i=1}^{n} p_i l_i \quad \text{(bits/signe)} \tag{14}$$

Une efficacité de 100% serait atteinte, pour une source sans mémoire, si la longueur de chaque mot était égale à la quantité d'information effective associée au signe correspondant, c'est-à-dire si $l_i = -\text{lb}\, p_i$. Dans ce cas, la longueur moyenne L s'identifierait à l'entropie H_0 donnée par la formule (7), de telle sorte que chaque bit d'encodage transporte un shannon d'information.

Ce résultat ne peut généralement être qu'approché car l_i est nécessairement un nombre entier. Shannon a montré qu'en encodant non les signes, pris individuellement, mais des blocs de m signes, il est toujours possible de s'approcher aussi près qu'on le veut d'une efficacité de 100%, même pour des sources avec mémoire, à la condition de choisir m assez grand. Ainsi,

l'information fournie par une source avec un débit effectif de H′ sh/s est théoriquement représentable (après codage) par un signal numérique ayant un débit apparent de H′ bit/s. Ceci constitue le *premier théorème fondamental* de la théorie de l'information. S'il indique l'existence de techniques permettant d'améliorer l'efficacité d'une source, il ne fournit pas de règles pour la construction du code.

Une méthode optimale a été proposée par HUFFMAN (v. CULLMANN, 1968). Elle consiste à coder les deux signes de plus faibles probabilités X_{n-1} et X_n à l'aide de mots binaires de longueurs identiques ne différant que par le dernier bit (voir figure 7). La procédure se répète ensuite après avoir remplacé dans la liste ces deux signes par un signe artificiel $X_{n-1,n}$ de probabilité cumulée $p_{n-1} + p_n$. On construit ainsi progressivement, en partant de la fin, un *arbre d'encodage* où à chaque branche est assigné un symbole d'encodage 0 ou 1 choisi de manière à assurer un décodage sans ambiguïté. Cette structure en arbre, impliquant qu'aucun mot du code ne puisse être le préfixe d'un autre mot du code, garantit un décodage instantané en ce sens que chaque signe peut être identifié instantanément dès

x_i	p_i	CODE	l_i	$p_i\ l_i$
x_1	0,4	0	1	0,40
x_2	0,15	1 1 0	3	0,45
x_3	0,15	1 0 0	3	0,45
x_4	0,1	1 0 1	3	0,30
x_5	0,1	1 1 1 0	4	0,40
x_6	0,06	1 1 1 1 0	5	0,30
x_7	0,02	1 1 1 1 1 0	6	0,12
x_8	0,02	1 1 1 1 1 1	6	0,12
			L = 2,54	

MESSAGE INITIAL

$x_2\ x_5\ x_1\ x_6\ x_8$

MESSAGE CODE

1 1 0 1 1 1 0 0 1 1 1 1 0 1 1 1 1 1

Figure 7. – Codage optimum de Huffman.

réception du dernier bit du mot-code correspondant. Dans l'exemple de la figure 7, on obtient une longueur moyenne d'encodage L = 2,54 bits/signe, alors que l'entropie de la source vaut H_0 = 2,485 sh/signe. La redondance de l'information, initialement de 17,2%, est ramenée par codage à 2,2%.

Une telle approche conduit dans les cas pratiques à des solutions généralement très compliquées, si bien que des techniques sous-optimales lui sont préférées.

Pour certains signaux analogiques (parole, télévision) une technique prometteuse est celle du codage différentiel. Elle consiste, pour l'essentiel, à établir une prédiction de la valeur présente du signal, basée sur l'acquis antérieur, et à n'encoder plus grossièrement et transmettre que la différence entre la valeur réelle et cette prédiction. Le récepteur reconstitue l'information de manière inverse. Lorsque l'erreur de prédiction est suffisamment faible, on peut même se contenter de ne transmettre que les paramètres, adaptés périodiquement, du système prédicteur. C'est ainsi que la parole, qui nécessite sans réduction de redondance un débit apparent de 64000 bits par seconde, peut être représentée après codage différentiel par un flux beaucoup plus faible de l'ordre de quelques milliers de bits par seconde (au prix toutefois d'une certaine dégradation de la qualité).

6. TRANSMISSION DE L'INFORMATION

Lors d'une transmission, les perturbations éventuelles (bruit) provoquent une modification du signal entraînant des erreurs d'interprétation (figure 8). Ceci introduit une certaine *ambiguïté* sur la nature de l'information transmise qui en réduit le débit effectif.

Avant transmission, la quantité d'information effective fournie en moyenne par une source X est l'entropie H(X) que l'on peut interpréter comme la mesure de l'incertitude initiale d'un observateur. L'ambiguïté de la transmission, due au bruit, correspond à l'incertitude finale que l'observateur a sur la nature de l'information après avoir pris connaissance du message reconstitué, avec de possibles erreurs, au récepteur Y. Cette ambiguïté moyenne – appelée «équivoque» par Shannon – s'exprime mathématiquement sous la forme d'une entropie de X conditionnelle à l'observation de Y: $H(X|Y) \leq H(X)$.

La quantité d'information effective I transmise en moyenne est alors simplement la différence entre l'incertitude initiale et l'incertitude finale

$$I = H(X) - H(X|Y) \tag{15}$$

Sans bruit, l'ambiguïté H(X|Y) est nulle et la quantité d'information transmise correspond à l'entropie de la source. Dans le cas limite d'une voie

interrompue, l'observation de Y ne nous apprend rien sur X et l'incertitude finale est égale à l'incertitude initiale. La quantité d'information transmise est donc nulle.

On appelle *capacité* C d'une voie de transmission le débit effectif maximum qu'elle autorise pour un régime de perturbation donné:

$$C = I_{max} \tag{16}$$

Dans le cas d'une voie transmettant des symboles binaires, dans laquelle la probabilité d'interprétation incorrecte d'un symbole est p, cette capacité vaut

$$C = 1 + p \operatorname{lb} p + (1-p)\operatorname{lb}(1-p) \qquad \text{(sh/bit)} \tag{17}$$

et est comprise entre zéro et un.

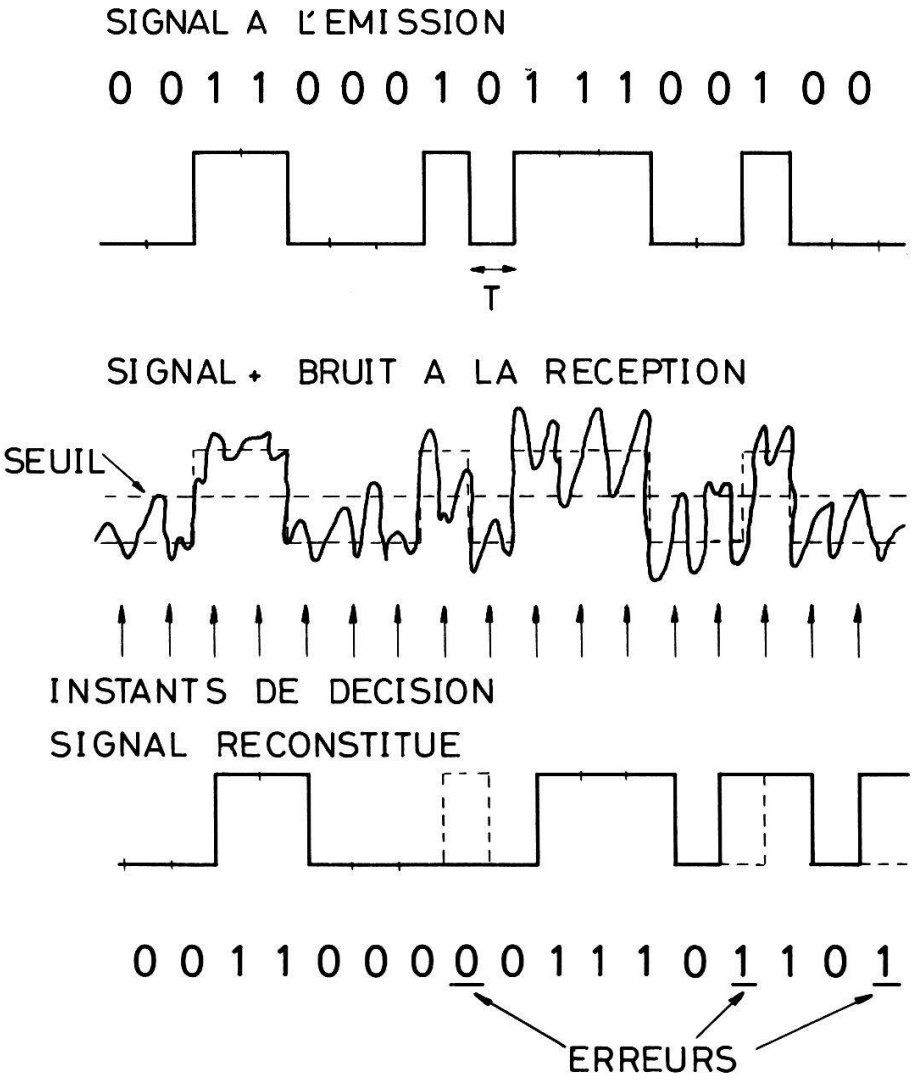

Figure 8. – Influence du bruit sur un système de transmission d'informations.

Pour une voie analogique, perturbée par du bruit gaussien, on obtient

$$C = B \cdot \mathrm{lb}\,(1 + P_s/P_b) \qquad (\mathrm{sh/s}) \qquad (18)$$

où B est la largeur de bande de la voie, P_s et P_b les puissances du signal et du bruit, respectivement. Cette formule célèbre établit une liaison entre les concepts d'énergie et d'information.

La capacité est une caractéristique très importante de la voie de transmission. On démontre en effet que tant que la transmission de l'information se fait sur une telle voie avec un débit inférieur ou égal à C, il est possible d'améliorer la sécurité de la liaison en réduisant la fréquence des erreurs d'interprétation à une valeur aussi faible que désirée moyennant l'emploi d'un codage redondant approprié. Une telle réduction ne peut pas être envisagée pour tout débit supérieur à C. Ce résultat est connu sous le nom de *deuxième théorème fondamental* de la théorie de l'information.

On peut comprendre intuitivement pourquoi la présence d'une redondance favorise la sécurité. Supposons la réception d'un télégramme contenant les mots THEORIX DE l'INFZRMATION. Il est certain que chaque lecteur rectifiera de lui-même en utilisant le contexte pour reconstruire le message correct.

7. Lutte contre les erreurs

Le deuxième théorème fondamental, proposé par Shannon, n'indique pas davantage que le premier comment il faut construire un code efficace (ici pour lutter contre les erreurs). La recherche de tels codes est en elle-même un domaine immense où se sont illustrés de nombreux chercheurs depuis 1948. Mentionnons, en particulier, la contribution importante du dernier conférencier de ce cours, le Professeur Marcel J.E. Golay, qui découvrit en 1949 déjà le seul code binaire parfait, capable de corriger toutes les erreurs simples, doubles ou triples dans un bloc de 23 bits.

Nous nous contenterons de présenter ici quelques exemples élémentaires de technique de codage permettant soit la détection, soit la correction d'erreurs.

La méthode la plus simple d'introduction d'une redondance limitée pour permettre la détection de certaines erreurs est l'adjonction, à chaque mot binaire représentant un signe de la source, d'un bit supplémentaire, dit de parité, choisi par exemple de telle manière que le nombre de symboles 0 dans le mot ainsi augmenté soit toujours pair. Un contrôle de parité est alors effectué à la réception qui décèle la présence de n'importe quelle combinaison impaire d'erreurs. Par exemple, si le message initial est représenté par le mot de 5 bits 10010, qui comprend trois symboles 0, on transmettra le mot 100100 dont le sixième bit est le symbole de parité.

Pour augmenter la probabilité de détection des erreurs, il faut augmenter la redondance. Un exemple classique est le code télégraphique à 7 bits qui est conçu de telle manière que chaque mot comprenne toujours quatre 1 et trois 0. Sur les $2^7 = 128$ combinaisons différentes de 7 bits, il en existe 35 qui ont la propriété ci-dessus. L'une de ces combinaisons est utilisée pour signaler en retour la détection d'une erreur à la réception et demander à l'émetteur la répétition du dernier message émis.

Lorsque le taux d'erreurs est trop élevé, exigeant une répétition trop fréquente des messages, et surtout lorsque le système de transmission ne fournit pas une voie de retour permettant de signaler la détection d'erreurs,

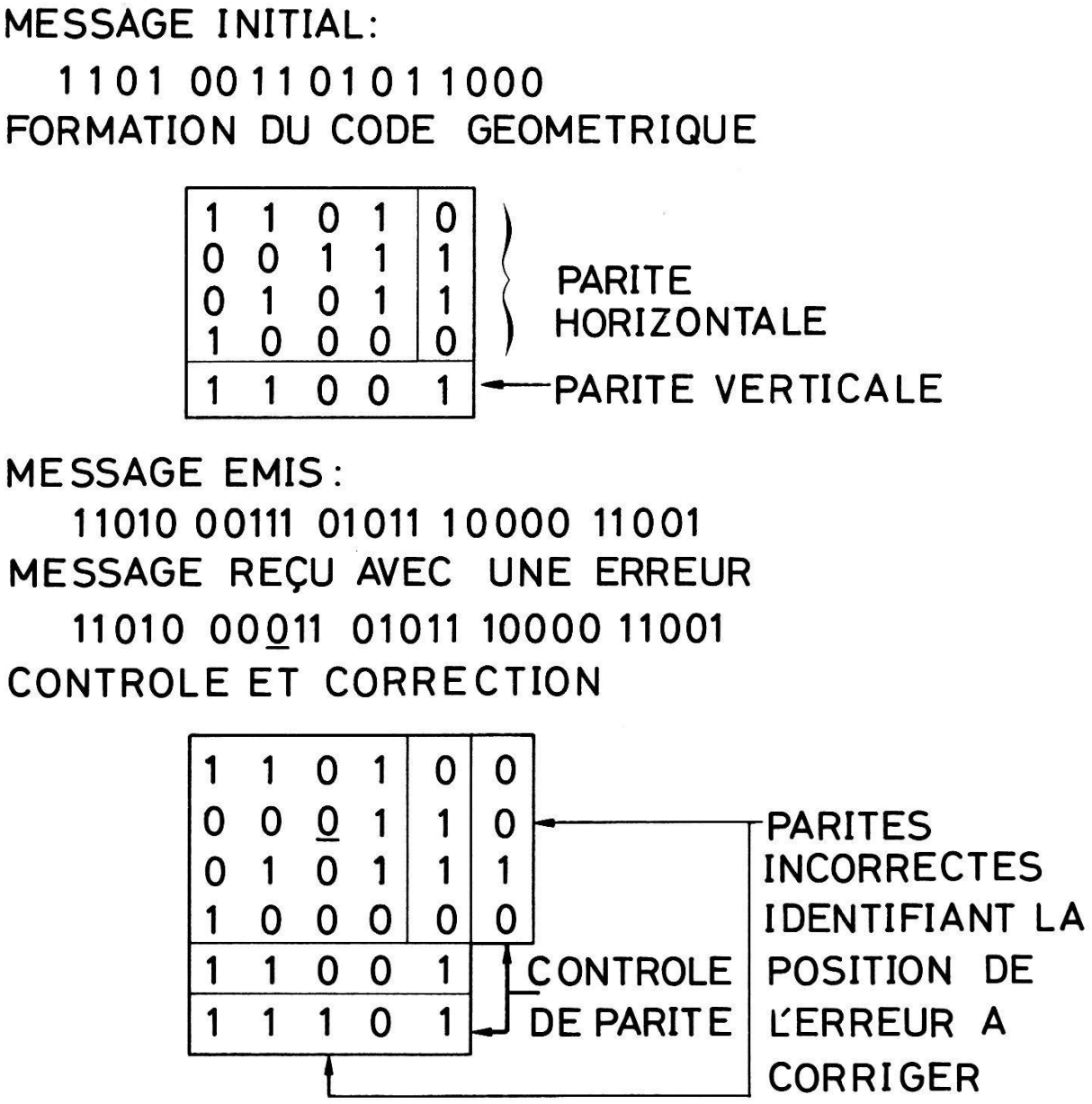

Figure 9. – Code géométrique corrigeant une erreur isolée.

la détection avec correction automatique de certaines erreurs devient le seul moyen d'améliorer la sécurité de la liaison.

Nous ne pouvons développer ici ce sujet. Afin d'illustrer toutefois le principe d'une correction d'erreur, on a représenté sur la figure 9 un exemple de code simple, corrigeant une erreur isolée, appelé code géométrique. Les symboles du message initial sont regroupés dans une matrice à laquelle on ajoute une ligne et une colonne contenant des symboles supplémentaires, assurant la parité horizontale et verticale (ici parité du nombre de zéro). La présence d'une erreur isolée dans le message fait échouer les deux contrôles de parité de la ligne et de la colonne auxquelles appartient l'erreur. Elle est donc détectée, localisée, et par conséquent corrigible. De plus, un tel code permet de détecter un grand nombre de configurations d'erreurs multiples.

8. Conclusion

Pour l'ingénieur, la théorie de l'information constitue une tentative en vue d'établir une base unique permettant de comparer les diverses méthodes de communication de l'information et de définir des critères permettant de mesurer leurs performances.

Les deux enseignements fondamentaux de la théorie développée par Shannon sont:

a) la preuve de l'existence de possibilités d'améliorer l'efficacité d'une source d'information redondante. Des solutions techniques ont été envisagées à cet égard pour des sources telles que texte, signal vocal, signal de télévision, documents facsimile, etc.;
b) la preuve de l'existence de possibilités d'améliorer la sécurité d'une transmission d'information en présence de perturbations – sans réduire excessivement le débit effectif d'information – grâce à l'introduction d'une redondance structurée qui permet de déceler, voire même de corriger, les erreurs.

La première décennie qui a suivi la publication des travaux de Shannon fut une période de recherche fondamentale très active en théorie de l'information. Les années soixante ont surtout permis à cette théorie de mûrir et de s'imposer. Mais c'est surtout au cours de la décennie actuelle que le développement d'applications concrètes s'intensifie. L'un des principaux facteurs de cette évolution réside dans l'étonnante croissance des moyens mis à disposition de l'ingénieur par la technologie micro-électronique (circuits intégrés, mémoires à grande capacité, microprocesseurs, etc.).

RÉFÉRENCES

CULLMANN, G. 1968. – Codage et transmission de l'information. Eyrolles, Paris.

SHANNON, C.E. 1948. – A mathematical theory of communication. *Bell System Technical Journal 27,* 379–423, 623–656.

WEAVER, W., SHANNON, C.E. 1975. – Théorie mathématique de la communication, traduction des articles de 1948, accompagnés d'une explication générale. Collection Les Classiques des sciences humaines. Retz-C.E.P.L., Paris.

WIENER, N. 1948 et 1961. – Cybernetics. MIT Press, Cambridge, USA.